IN THE SPECIFICATION:

Please replace paragraph [0006] at page 2 with the following paragraph:

[0006] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 33945 protein or polypeptide, *e.g.*, a biologically active portion of the 33945 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides isolated 33945 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1[[,]] or SEQ ID NO:3 ~~or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____~~. In still other embodiments, the invention provides nucleic acid molecules that are sufficiently or substantially identical (*e.g.*, naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1[[,]] or SEQ ID NO:3 ~~or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____~~. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1[[,]] or SEQ ID NO:3 ~~or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC Accession Number ____~~, wherein the nucleic acid encodes a full length 33945 protein or an active fragment thereof.

Please replace paragraph [0011] at page 3 with the following paragraph:

[0011] In other embodiments, the invention provides 33945 polypeptides, *e.g.*, a 33945 polypeptide having the amino acid sequence shown in SEQ ID NO:2 ~~or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____~~; an amino acid sequence that is sufficiently or substantially identical to the amino acid sequence shown in SEQ ID NO:2 ~~or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____~~; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3 ~~or the nucleotide sequence encoded by the cDNA insert of the plasmid deposited with ATCC Accession Number ____~~, wherein the nucleic acid encodes a full length 33945 protein or an active fragment thereof.

Please delete paragraph [0034] at page 6.

Please replace paragraph [0022] at page 5 with the following paragraph:

[0022] Human 33945 contains the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 420 ~~and http://www.pse.edu/general/software/packages/pfam/pfam.html~~ or the Pfam website maintained in several locations, *e.g.* by the Sanger Institute (pfam.sanger.ac.uk), Washington University (pfam.wustl.edu), the Karolinska Institute (pfam.cgr.kr.se) or Institut de la National Recherche Agronomique (pfam.jouy.inra.fr)):

Please replace paragraph [0037] at page 7 with the following paragraph:

[0037] Glycosyltransferases can have two types of topology. Glycosyltransferases of the Golgi do not possess an obvious sequence homology which would suggest a common Golgi retention signal. However, they are all membrane proteins and share type II topology. The Type II topology of glycosyltransferases (also shared by group 2 glycosyltransferases), consists essentially of an amino terminal cytoplasmic tail, a signal anchor transmembrane domain, a stem region, and a large luminal catalytic domain. The membrane-spanning domain and its flanking regions contain necessary and sufficient information for Golgi retention of these enzymes (Jaskiewicz (1997) *Acta Biochim Pol* 44:173-9). Endoplasmic reticulum (ER) localized glycosyltransferases can have either a type II topology, like the Golgi glycosyltransferases, or a type I topology, e.g., the N-terminus and catalytic domain inside the ER (Kapitonov et al. (1999) *Glycobiology* 9:961-78). The 33945 protein is homologous to ProDom family PD003162 ("N-acetylgalactosaminyltransferase Transferase Polypeptide Acetylgalactosaminyltransferase UDP-GalNac:polypeptide Glycosyltransferase Protein-UDP-protein- UDP N-;" SEQ ID NO:7; ProDomain Release 2000.1; for ProDom information, refer to Institut National de la Recherche Agronomique (INRA)/ Central National de la Recherche Scientifique (CNRA), Toulouse, France ~~http://www.toulouse.inra.fr/prodom.html~~). An alignment of 33945 with this consensus sequence shows 61% identity in the region at about amino acid residues 287 to 443 of SEQ ID NO:2. The 33945 protein shares 57.9% identity with mouse polypeptide GalNac transferase-T4, another glycosyltransferase (type 2) family member (Accession number 2121220 in GenPept, SEQ ID NO:8) as calculated from a matrix made by matblas from blosum62.iij.

Please replace paragraph [0039] at page 7 with the following paragraph:

[0039] As used herein, the term "glycosyltransferase domain" includes an amino acid sequence of about 100 to 250 amino acid residues in length and having a bit score for the alignment of the sequence to the glycosyltransferase domain (HMM) of at least 40. Preferably, a glycosyltransferase domain includes at least about 120 to 220 amino acids, more preferably about 140 to 200 amino acid residues, or about 160 to 190 amino acids and has a bit score for the alignment of the sequence to the glycosyltransferase domain (HMM) of at least 50, 60, 80 or greater. Preferably a glycosyltransferase domain mediates the transfer sugar from UDP-glucose, UDP-N-acetyl-galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids. Glycosyltransferase domains (HMM) have been assigned numerous PFAM Accession Numbers, including PF00534 (group 1) and PF00535 (group 2) (~~http://pfam.wustl.edu/~~ see Pfam information at Washington University in St. Louis, MO). An alignment of the glycosyltransferase domain (amino acids 139 to 322 of SEQ ID NO:2) of human 33945 with a consensus amino acid sequence (group 2 glycosyltransferases, SEQ ID NO:4) derived from a hidden Markov model yields a bit score of 85.1.

Please replace paragraph [0042] at page 8 with the following paragraph:

[0042] As used herein, the term "ricin domain" includes a protein or polypeptide which is capable of recognizing, e.g., binding, a sugar molecule and has an amino acid sequence of about 80 to 200 amino acid residues in length and having a bit score for the alignment of the sequence to the ricin domain based on SMART of at least 40 (see SMART information at European Molecular Biology Laboratory (EMBL), Heidelberg, Germany ~~http://smart.embl-heidelberg.de/~~). Ricin domains are typically present in many carbohydrate binding proteins, e.g., plant and bacterial AB toxins, glycosidases and proteases. This domain, also known as the ricin B lectin domain, can be present in one or more copies and has been shown in some instances to bind to simple sugars, such as galactose and lactose. Preferably, a ricin domain includes at least about 100 to 180 amino acids, more preferably about 120 to 160 amino acid residues, or about 130 to150 amino acids and has a bit score for the alignment of the sequence to the ricin domain (HMM) of at least 50, 60, 70 or greater. An alignment of the ricin domain (amino acids 441 to 577 of SEQ ID NO:2) of human 33945 with a consensus amino acid sequence (Ricin_B_lectin, Pfam) derived from a hidden Markov model yields a bit score of 18.7 and an alignment of the ricin domain (amino acids 445 to 577 of SEQ ID NO:2) of human 33945 with a consensus amino

acid sequence (Ricin_3, SMART) derived from modular architecture analysis yields a bit score of 73.3.

Please replace paragraphs [0044] and [0045] at page 9 with the following paragraphs:

[0044] To identify the presence of a "glycosyltransferase" domain or a "ricin" domain in a 33945 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the Pfam database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (~~http://www.sanger.ac.uk/Software/Pfam/HMM_search~~ see the Pfam website maintained in several locations, *e.g.* by the Sanger Institute (pfam.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a "glycosyltransferase domain" domain in the amino acid sequence of human 33945 at about residues 139 to 322 of SEQ ID NO:2 and a "ricin domain" in the amino acid sequence of human 33945 at about residues 441 to 577 of SEQ ID NO:2.

[0045] The presence of a "ricin" domain in a 33945 protein sequence can also be identified using a SMART database (Simple Modular Architecture Research Tool, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany ~~http://smart.embl-heidelberg.de/~~) of HMMs as described in Schultz *et al.* (1998), *Proc. Natl. Acad. Sci. USA* 95:5857 and Schultz *et al.* (2000) *Nucl. Acids Res* 28:231. The database contains domains identified by profiling with the hidden Markov models of the HMMer2 search program (R. Durbin *et al.* (1998) *Biological sequence analysis: probabilistic models of proteins and nucleic acids.* Cambridge University Press.; see HMMER information at Washington University in St. Louis, MO ~~http://hmmer.wustl.edu/~~). The database also is extensively annotated and monitored by experts to enhance accuracy. A search was performed against the HMM database resulting in the

identification of a "glycosyltransferase" domain in the amino acid sequence of human 33945 at about residues 445 to 577 of SEQ ID NO:2.

Please replace paragraph [0081] at page 24 with the following paragraph:

[0081] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at ~~http://www.geg.com~~ the bioinformatics page of the website maintained by Accelrys, Inc., San Diego, CA, USA), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package ~~(available at http://www.geg.com)~~, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

Please replace paragraph [0083] at page 24 with the following paragraph:

[0083] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 33945 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 33945 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and

NBLAST) can be used~~. See http://www.ncbi.nlm.nih.gov~~ (accessible at the website maintained by National Center for Biotechnology Information, Bethesda, MD, USA).